

Казанский государственный университет

Исторический факультет

Кафедра современной отечественной истории

Н.А.Федорова

ЗАДАЧНИК

**для студентов отделения «История»
по курсу «Применение математических методов
в историческом исследовании»**

Казань

2007

ВВЕДЕНИЕ

Курс «Применение математических методов в историческом исследовании» читается студентам-историкам дневного отделения на втором курсе и состоит из лекций и лабораторных практических занятий. Студенты заочного отделения изучают этот предмет на третьем курсе только лекционным путем. Данная учебная дисциплина призвана научить будущего специалиста навыкам работы со структурированными источниками на основе методов математической статистики. Современные информационные технологии позволяют автоматизировать процесс реализации используемых методик, однако для успеха необходимо, чтобы студент разбирался в сути метода и мог грамотно поставить исследовательскую задачу. Для этого в ходе лабораторного практикума учащимся предлагаются задания, выполнить которые они должны без применения сложной вычислительной техники, ограничиваясь письменными принадлежностями, линейкой, калькулятором (опыт показывает плачевные результаты современной школьной подготовки – молодежь не способна удержать в памяти таблицу умножения и не владеет устным счетом).

На следующем этапе обучения студенты знакомятся с основами исторической информатики, где в качестве закрепления и совершенствования полученных знаний они овладевают компьютерными методами анализа данных.

В результате освоения материала студент должен уметь работать с табличными данными, представлять их в графической форме, применять математико-статистические методы анализа, разбираться в их методологических основах, а также познакомиться с современной специальной литературой, содержащей опыт применения количественных методик в исторических исследованиях.

Предлагаемый «Задачник» составлен в соответствии с Программой и курсом лекций Н.А.Федоровой «Математические методы в историческом исследовании» (Казань, изд-во Форт Диалог, 1996). Он не охватывает всех математико-статистических методов, изучаемых в курсе. В него вошли наиболее часто встречаемые методики. Решение подобных примеров должно помочь студентам избавиться от наиболее типичных ошибок, до-

пускаемых при работе со статистическими данными. «Задачник» рассчитан на самостоятельную работу в качестве вспомогательного учебного пособия. В тексте почти нет математических формул, их можно найти в специальной литературе или в опубликованном курсе лекций Н.А.Федоровой (см. выше). Задачи снабжены примерными рассуждениями, раскрывающими алгоритм решения различных типов задач. Текст выдержан в стиле беседы, разговора Преподавателя со Студентом. Думается, что самостоятельная работа с «Задачником» поможет учащимся в освоении курса, как в теоретическом, так и в практическом плане, в выработке навыков самостоятельной исследовательской работы.

МЕТОД ГРУППИРОВКИ

Любое исследование начинается с группировки разрозненных документов, фактов, характеристик того или иного процесса. В одну группу объединяются чем-то похожие единицы. Их различия внутри группы должны быть менее существенными для исследования, чем различия между группами.

Как производится группировка?

Разберем простой пример: перед нами геометрические фигуры –



Предположим, что все они разного цвета, разного размера и представлены в исходной изучаемой совокупности каждая в разном количестве. Самое простое – сгруппировать эти предметы по форме. В первую группу войдут все кружочки, во вторую – треугольники, в третью группу – квадраты. Можно сгруппировать эти объекты по схожести цвета (групп будет столько, во сколько цветов окрашены фигуры). Можно вычислить площадь объектов и в соответствии с этим показателем разбить исходную совокупность на группы. Можно выделить две группы – «угловатую» и «круглую» и т.д.

От чего зависит выбор способа группировки? – От цели работы. В зависимости от цели исследования мы и выбираем признак, который будет

положен в основу группировки. Этот признак называется базовым или группировочным. Он может быть качественным и количественным. В чем разница?

«Рост» – какой признак, качественный или количественный? Все зависит от того, чем выражен признак. Если он выражен словом, то его относят к разряду качественных, а если числом – количественных. Как видишь, деление довольно условно, т.к. один и тот же признак может быть и качественным, и количественным. Если мы скажем: «В комнату вошел высокий старик», то признак «рост» рассматривается как качественный (поскольку выражен словом «высокий»). Если же эта фраза примет иное звучание: «В комнату вошел старик, рост которого превышал 180 сантиметров», то в этом случае «рост» – количественный признак (т.к. выражен числом 180).

Следовательно, качественные и количественные признаки различаются способом своего выражения.

Проверь себя:

В каком из приведенных высказываний присутствует качественный, а в каком – количественный признак?

1. Крестьянские хозяйства бывают бедняцкими, середняцкими и кулацкими.
2. По форме собственности мелкая промышленность делилась на государственную, кооперативную и частную.
3. Посевная площадь на одно хозяйство в 1916 г. в среднем составляла 3,9 десятин, а в 1922 г. упала до 2, 12 десятин.

Подсказка: Для начала выдели признак, о котором идет речь, затем посмотри – чем он выражен. Сразу видно, что в первом и втором примерах нет цифр, только буквы, а значит и признаки там качественные. В третьем примере речь идет о посевной площади, размер которой указан в цифрах. Уверенно отвечаем, что признак здесь использован количественный.

Усложним задачу, обратимся к таблице.

Таблица 1

Плотность населения кантонов Татареспублики в 1920 г.*

Кантоны	Плотность населения	Кантоны	Плотность населения
Агрызский	14, 25	Мензелинский	39, 64
Арский	57, 66	Свияжский	59, 17
Бугульминский	37, 20	Спасский	44, 85
Буинский	61,99	Тетюшский	72, 31
Елабужский	46, 62	Челнинский	47,51
Лаишевский	47,92	Чистопольский	46, 52
Мамадышский	47, 97	Итого по ТАССР	48, 76

* Труд и хозяйство, 1924, № 4, с. 6.

Таблица 2

Обеспеченность сельского населения скотом в 1920 г.*

Скот	Количество голов
Лошади	486 507
Крупный рогатый скот	549 008
Свиньи	192 720
Мелкий скот	1 222 780

* Труд и хозяйство, 1924, № 4, с. 7-8.

Подсказка: Для начала выделим признак, о котором идет речь. В Табл.1 приведены данные о плотности населения. Признак выражен числами (указано количество проживающих на 1 квадратную версту), следовательно, он количественный. В Табл. 2 дано распределение скота в сельской местности. Таблица содержит и количественные показатели, но что здесь является признаком? Посмотри заголовок, название таблицы и станет ясно, что речь идет о разных видах скота – лошадях, свиньях, крупном рогатом скоте и т.д. Виды эти выражены словами, значит признак качественный. Чем же тогда являются числа в этой таблице? – Показателями частоты проявлений конкретных значений признака в изучаемой совокупности. Мы

видим, что лошадей было 486 507. Это означает, что признак «скот» в значении «лошадь» встречается в исходной совокупности объектов 486 507 раз, или – частота (Р) данного значения признака равна 486 507.

Проверь себя: Чему равна частота признака в значении «мелкий скот»? Если твой ответ 1 222 780, то ты прав.

Идем дальше. По базовому (он же группировочный) признаку определяется вид группировки. Положенный в основу качественный признак формирует типологическую группировку (см. Табл. 2), количественный признак – структурную (см. Табл. 1).

А теперь поработай самостоятельно.

1. Выдели количественные признаки:

Пол; возраст; национальность; место рождения; семейное положение; количество детей; адрес, согласно прописке.

2. Определи вид группировки.

Таблица 3

Землеобеспеченность крестьянских хозяйств (в десятинах в расчете на 1 едока) в 1923 г.*

Земл. в дес.	0- 0,25	0,25 -0,5	0,5 – 0,75	0,75 – 1,0	1,0 – 1,5	1,5 – 2,0	2,0 – 2,5	2,5 – 3,0	Более 3,0
Кол- во хоз-в	1223	4400	16946	45000	176343	146493	50638	14694	10874

* Труд и хозяйство, 1923, № 6, с.50

Таблица 4

Средняя урожайность в Татарской Республике в 1927 г.*

С/х культуры	Рожь	Пшеница	Овес	Греча	Картофель
Урожай (в пуд/дес.)	50,8	18,0	27,8	24,7	384,0

* Тарханов О. Десятилетие Советского Татарстана (1920-1930). – Казань: Татиздат, 1930. – С.23.

Таблица 5

Распределение ходатайств лишенцев о восстановлении в правах*

Основание для восстановления	Количество ходатайств
Пролетарское происхождение	5
Незначительный объем торговли и непродолжительный период занятия ею	10
Вынужденный характер занятия торговлей	8
Труд по найму, добросовестное отношение к труду	30
Лояльность к власти, прошлые заслуги перед властью	17
Несправедливость решения о лишении избирательных прав	27
Беспокойство за судьбу детей	3

*Социальный портрет лишенца (на материалах Урала). – Екатеринбург, 1996. – С.223.

Таблица 6

Национальный состав сельского населения ТАССР в 1920 г.*

Националь- ность	татары	рус- ские	чува- ши	морд- ва	вотя- ки	мари	Про- чие
Кол-во жите- лей	145250 6	980292	141795	39017	23198	1301 5	2684

* Труд и хозяйство, 1925, № 11, с.25.

3. Что необходимо сделать, чтобы преобразовать структурные группировки в типологические? Попробуй совершить это превращение с приведенными здесь группировками.

Таблица 7

Распределение малых предприятий по прибыли и оборачиваемости капитала в середине 90-х годов XX в.*

Продолжительность оборота капитала (в днях)	Число малых предприятий	Средняя прибыль (млн. руб.)
40 – 50	6	14, 6
51 – 70	8	12, 9
71 – 100	6	7, 4

* Таблица учебная, данные условные.

Таблица 8

Распределение крестьянских хозяйств по обеспеченности землей и скотом в 1923 г.*

Землеобеспеченность (в десятинах)	Количество голов домашнего скота (на хозяйство)			
	0	1 – 2	3 – 4	5 и более
0 – 1	28980	36387	2128	74
1 – 2	134429	216428	21242	1375
2,5 – и более	9947	12966	2338	317

* Труд и хозяйство, 1923, № 7, с.53

Что же это за группировка, в которой присутствуют два признака (см. Табл. 7, 8)? – Аналитическая. Она позволяет выявить связь между признаками и дать заключение о характере, прямая или обратная, этой связи.

Здесь необходимо обращать внимание на два момента. Во-первых, прочитав название таблицы или самостоятельно группируя признаки, хорошенько подумай – возможна ли в принципе какая-либо связь между ними, могут ли они оказывать влияние друг на друга? Например, национальность не зависит от места рождения, им не определяется; возраст не зависит от пола и наоборот ... Во-вторых, посмотри как распределились в таблице переменные значения признаков, не концентрируются ли они по диа-

гонали? Если такое явление имеет место – определи направление диагонали и после этого можешь делать выводы. Забыл, какие выводы можно сделать – загляни в курс лекций, либо в конспекты, либо в книгу.

Проверь себя:

Какие признаки из перечисленных можно свести в аналитическую группировку?

Пол; возраст; национальность; место рождения; семейное состояние; количество детей; наличие гражданских прав; имущественное положение; площадь землевладения; главное занятие; наличие в хозяйстве лошади, коровы, мелкого скота; использование наемной рабочей силы, участие в политических организациях, служба в армии.

А теперь поработай самостоятельно:

Найди ошибки в построении и оформлении приведенных ниже таблиц.

Таблица 9

Кол-во детей в семье	нет	1-2	2-3	3-5
Кол-во семей	1287	2294	2808	1146

Распределение работников предприятия по месту рождения и полу

Место рождения	Город	Сельская местность
Мужчины	17654	354709
Женщины	19872	395126

возраст	15-17	18-21	22-26	26-28	более 30
мужчины	12	6%	72,3	184	18
женщины	61.7	22	3,21	17%	190

* * *

Если в группировку включен хронологический показатель, перед нами не просто группировка, а динамический ряд. Он требует к себе особого внимания, а значит и особых методов анализа.

Динамические ряды могут быть моментными и интервальными.

Проверь себя:

Какие из приведенных таблиц относятся к моментным, а какие – к интервальным динамическим рядам?

Таблица 10

Изменение посевных площадей в Тираспольском уезде.*

Год	1914	1915	1916	1917	1918	1919
Общая площ. посева	474.59	453.56	439.40	397.11	422.0	331.50
Площ. посева зерна	465.08	456.73	438.46	388.41	393.13	316.20

* Кикуш Н.Е. Экономическое развитие доколхозного села Молдавской АССР. – Кишинев, 1989. – С.56.

Таблица 11

Ожидаемая продолжительность жизни в СССР в 1959-1979 гг.*

годы	1959- 1960	1961- 1969	1970- 1975	1976- 1978	1979
Муж.	64.4	65.7	64.5	62.9	62.5
Жен.	71.7	73.8	73.5	72.6	72.5

* Население Советского Союза. – М., 1993. – С.95

Таблица 12

Уровень смертности в ТАССР в 1921/1922 гг.
(по данным на 1 число каждого месяца)*

Время	X1. 1921	X11. 1921	1. 1922	1 11 1922	111. 1922	1У. 1922	У. 1922
Всего Умерло	75	3925	5729	13444	19848	17845	9401
Из них детей	23	921	3902	9655	15427	8736	3287

*НАРТ, ф.4470, оп.1,д.39,Л.49.

Таблица 13

Валовый сбор хлебов в Татарстане.*

Валовый сбор хлебов	1909 – 1914	1918 – 1920	1921 – 1926
В тыс.пудов	112 403	87 493	53 044
В %	100	78,8	47,3

* Шайдуллин Р.В. Крестьянство Татарстана: экономический и общественно-политический аспекты (1920 – 1929 гг.). – Казань: изд-во «Фэн» АН РТ, 2004. – С.59.

Структурную группировку построить легче, если предварительно базовый признак представить в виде ранжированного ряда, т.е. таблицы, в которой значения признака систематизированы в возрастающем или убывающем порядке.

Проверь себя:

Какая таблица из приведенных ниже является структурной группировкой, а какая – ранжированным рядом? Чем отличается ранжированный ряд от структурной группировки?

Таблица 14

Обеспеченность крестьянских хозяйств крупным рогатым скотом*

А)

Кол-во голов к.р.с.	0	1	2	3	4	5	6	7
Кол-во хоз-в	4	7	6	4	1	2	0	1

Б)

Кол-во голов к.р.с.	0	1 – 3	4 – 6	7 и более
Кол-во хоз-в	4	17	3	1

* Таблица учебная. Данные условные.

ГРАФИЧЕСКИЙ МЕТОД

Табличные данные можно представить в виде графиков. Зачем? Во-первых, график помогает выявить общую тенденцию развития анализируемого признака, процесса или явления. Во-вторых, график придает эмоциональную окраску сухому изложению материала. Наглядность повышает эффективность восприятия, заставляет обращать внимание на то, что ускользало из восприятия материала представленного в иной форме.

Выбор формы графика зависит от цели исследования, от характера исходных данных, но в большей степени от цели применения данного метода. Надо решить – зачем нам нужен график? Что мы хотим продемонстрировать с его помощью – закономерность? Тенденцию? Соотношение частей в целом? Какие-то определенные показатели и величины?

Когда ты ответил себе на эти вопросы, можно приступить к построению (кстати, графики именно строят, а не рисуют) графика.

Начнем с гистограммы распределения. Вспомним, что она представляет собой столбиковую диаграмму и применяется для интервального ряда, т.е. когда значения изучаемого признака выражены интервалами. Далее, нельзя забывать, что значения признака всегда размещаются на линии ОХ (ось абсцисс), а значения частот (Р) – на линии ОУ (ось ординат).

Проверь себя – соблюдаешь ли ты масштаб при нанесении координат на оси ОХ и ОУ? Не случилось ли так, что интервалы значений группировочного признака в таблице разные по величине, а расстояния, откладываемые для них на оси абсцисс одинаковые? Допустим, что для интервала «6-8» ты наметил в системе координат 1 см, а сколько см. надо отложить в той же системе для интервала «8-14»?

Если ты помнишь одно из главных правил построения графика, согласно которому для частот больших интервалов на гистограмме распределения берется меньший масштаб (и наоборот), то начинай строить графики.

А теперь поработай самостоятельно с этими таблицами:

Таблица 15

Распределение неместных уроженцев Среднего Поволжья по возрасту*

Возраст	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70 и старше
Муж.	67156	135256	150104	105087	82956	49082	27200	12153
Жен.	68609	155831	253090	199154	154639	117594	75917	38093

* ЦСУ СССР. Всесоюзная перепись населения 1926 г. – Т.37. – С. 150–173.

Таблица 16

Распределение городов Российской Федерации по численности постоянного населения в 1992 г.*

Жит. в тыс. чел.	Менее 4.9	5-9.9	10-19.9	20-49.9	50-99.9	100-249.9	250-499.9	500-999.9	1млн и более
Кол-во городов	24	80	251	372	167	91	46	21	12

* Российская Федерация в 1992 г. Статистический ежегодник. – М., 1993. – С.98.

Таблица 17

Удельный вес грамотного татарского населения
по данным Переписи 1926 г.*

Возраст	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-90	90-99
Муж	4.1	60.5	70.8	60.6	41.9	31.7	25.4	39,1	15.5
Жен	2.7	46.2	43.0	24.8	16.8	14.7	12.1	17,1	8.9

* ЦСУ СССР. Всесоюзная перепись населения 1926 г. – Т.3. – С. 206-263.

Ты не забыл про плотность (f)? Она помогает, когда все интервалы значений признака неравны по величине. Подсказка: Плотность – величина равная отношению частоты признака к величине соответствующего интервала или $f = P_i/h_i$.

В этом случае высота столбика на графике будет равна величине плотности соответствующего интервала. Только не надо (нельзя, недопустимо) смешивать частоту и плотность в одном графике и высоту части столбиков определять по плотности, а части – по частоте.

По полученным гистограммам легко построить полигон распределения, соединив ломаной линией середины верхних оснований столбиков. Не забудь, что линия полигона начинается всегда из нулевой точки и заканчиваться должна нулевым значением, образуя замкнутую ломаную линию. А еще полигон строят для повышения исторической достоверности графика. Согласись, что в жизни общества, в истории невозможны столь резкие и четкие изменения, какие мы наблюдаем на гистограмме, сравнивая ее столбики. Ты уже построил гистограммы распределения по материалам таблиц 15 – 17. Теперь на этих же графиках получи полигоны.

Кроме того, полигон строят, если признак выражен дискретными величинами без интервальной разбивки. Вот, попробуй для тренировки получить полигон распределения по этим данным:

Таблица 18

Распределение рабочих-делегатов XXIX отчетно-выборной комсомольской конференции 1979 г. по возрасту.*

Коллет	18	19	20	21	22	23	24	25	26	27
Колраб	3	2	12	7	10	2	4	18	10	15

* Таблица учебная. Данные условные.

А теперь обратимся к кумуляте.

Подсказка: Это вид линейного графика, представляющий плавную кривую, устремленную вверх, поскольку на оси ординат откладываются значения не просто, а накопленных частот. Следовательно, чтобы построить кумуляту надо предварительно получить ряд накопленных частот. Можно его оформить в виде дополнительной строчки к таблице. Он получается путем последовательного сложения частоты каждого переменного значения признака с суммой предыдущих частот. Это не так сложно, как кажется. Вот посмотри:

X	0-10	10-20	20-30
P	16,7	41,2	59,0

Это фрагмент таблицы, по которому мы поучимся формировать ряд накопленных частот.

Итак, первая частота (P_1) – 16,7. У нее нет суммы предыдущих частот, она – первая. Оставляем ее на своем месте, а затем складываем первую частоту со второй (P_1+P_2) – 16,7+41,2. Получаем результат – 57,9. Это и есть вторая из накопленных частот. После этого к сумме двух частот, первой и накопленной, прибавляем третью (P_3) – 59,0. Получаем 116,9. Таким образом, ряд накопленных частот выглядит так: 16,7; 57,9; 116,9. Можно его для удобства дальнейших действий пристроить к нашему фрагменту учебной таблицы.

X	0-10	10-20	20-30
P	16,7	41,2	59,0
ΣP	16,7	57,9	116,9

Теперь понятно, каким путем получают ряд накопленных частот? Если нет – читай все сначала, если понял – строй график, кумуляту. В качест-

ве исходных данных используй данные предыдущих таблиц, по которым ты строил гистограммы и полигоны распределения. Кумулята пригодна как для случаев интервальной разбивки значений признака, так и для дискретных значений.

А ты заметил, что все таблицы, по которым мы строили графики, относятся к структурному виду группировок? Как быть с типологическими группировками?

Для них более подходит круговая диаграмма. Она представляет собой круг, из центра которого выходят векторы, делящие площадь круга на сегменты, соответствующие частотам рассматриваемых значений признака. Это очень удобно, т.к. качественный признак отложить на оси абсцисс каким-то делением прямой линии можно, но выглядеть это будет как-то несуразно. Вот, сам сравни:

Рис.1. Рис.2.

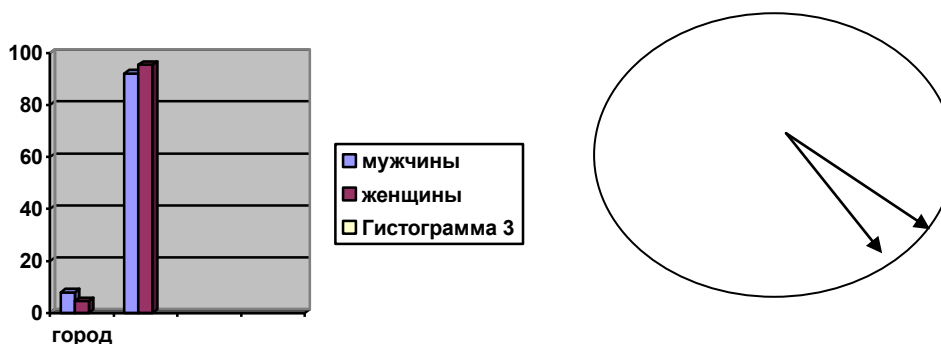


Рис.1 и Рис.2. Рабочие N-ского завода по месту рождения (город, сельские местности).

Круговая диаграмма позволяет представить наглядно соотношение значений признака (в нашем примере признак – место рождения, его значения – город, сельская местность) в изучаемой совокупности объектов.

А теперь поработай самостоятельно

Построй диаграммы по данным этих таблиц:

Таблица 19

Поголовье скота в крестьянских хозяйствах Одесской губернии в 1921 г.*

Скот	Лошади	Волы	Коровы	Овцы
Кол-во голов	8580	6881	10462	6635

* Таблица учебная. Данные условные.

Таблица 20

Распределение землевладельцев Казанской губернии по сословной принадлежности.*

Сословие	дворяне	купцы	мещане	крестьяне
Число землевл.	705	225	149	1687

* 20 лет Татарской АССР (1920 – 1940). – Казань: Татгосиздат, 1940. – С.102.

Таблица 21

Социальный состав деревни Казанской губернии (по данным земской статистики).*

Социальный состав	Количество в %
Безхозяйственные и отсутствующие	24,0 %
Бедняки	33,5 %
Середняки	29,7 %
Капиталистический тип	12,8 %

* 20 лет Татарской АССР (1920 – 1940). – Казань: Татгосиздат, 1940. – С.102.

Для динамических рядов тоже строится график. Напоминаю, что на оси абсцисс откладывается шкала времени, а на оси ординат – значения уровней ряда (та не забыл, что так называются показатели динамики?) В установленной системе координат точками отмечаются показатели динамического ряда. Эти точки соединяются прямыми и в результате возникает ломаная линия. В отличие от полигона распределения она никогда не образует замкнутого пространства. Даже, если по данным источника изучаемое явление начинается с нуля, с нулевой точки, график динамического ряда не может закончиться на нулевой отметке.

А теперь поработай самостоятельно

1. Построй графики для динамических рядов.
2. Вспомни, что такое тренд. Если забыл, загляни в свои конспекты или в курс лекций Н.А.Федоровой «Математические методы в историческом исследовании» на страницу 36 или освежи свою память как-то иначе, но определи тренд по данным примерам:

Таблица 22

Поголовье скота в крестьянских хозяйствах*

год	лошади	коровы	овцы	свиньи
1924	72599	57474	264068	64226
1925	69427	55615	172727	37024
1926	75230	58997	154367	29152
1927	86493	54115	147900	48856
1928	96205	56120	151997	70959
1929	91556	52959	127290	19881

*Кикуш Н.Е. Экономическое развитие доколхозного села Молдавской АССР. – Кишинев, 1989. –С.104

Таблица 23

Способы самоубийств в Казани*

Методы	1995	1996	1997	1998	1999
Повешение	488	490	414	434	405
Падение с высоты	59	55	64	70	64
Рельсовые травмы	10	12	8	4	5
Колото-резанные раны	27	21	20	17	19
Огнестрельные травмы	12	11	13	9	10
Отравление	40	43	38	22	27
Редкие виды самоубийств	1	4	9	6	6
всего	637	546	566	569	536

* Забусов Ю.Г. и др. Мониторинг и судебно-медицинские аспекты насильственной и внезапной смерти в Республике Татарстан. – Казань, 2000. – С. 62.

СРЕДНИЕ ВЕЛИЧИНЫ

Одной из главных задач научного исследования было и остается выявление закономерностей. Метод средних величин, характеристики всех единиц изучаемой совокупности объектов и игнорируя их индивидуальные особенности, устанавливает то общее, что в конкретных обстоятельствах места и времени выступает равнодействующей всех причин и сил, влияющих на изучаемое явление. Средняя является специфической формой выражения содержания общего закона, который выступает в виде тенденции.

А теперь будем учиться определять средние величины, а главное – научимся ориентироваться в каких случаях, какая средняя величина нам даст наилучший результат.

Самая распространенная – средняя арифметическая. Ее использование часто даже не оговаривают в тексте исследования: если автор пишет о средних показателях, подразумевается средняя арифметическая. Она определяется путем отношения суммы всех значений признака к общему числу наблюдений.

Проверь себя: У студента А – 15 компакт-дисков, у студента Б – 22 компакт-диска, у студента В – 48 компакт-диска с записями классической музыки. Определи среднее количество дисков, имеющихся у данной группы студентов.

Что делаем? Складываем общее количество компакт-дисков $(15+22+48)$ и делим на количество студентов (3).

$$\bar{X} = (15+22+48)/3 = 28,33$$

Мы получили величину (28,33), демонстрирующую, сколько компакт-дисков было бы у каждого из 3-х студентов, если бы им вздумалось поделить их поровну. Сразу обратим внимание на то, что ни у кого в рассматриваемой группе не было такого количества дисков, какое показала средняя арифметическая. И еще – явно одним диском пришлось бы пожертвовать ради справедливости, поскольку 0,33 диска – это разбитый диск. Оба сделанных замечания выводят нас на важную характеристику средней арифметической, которую важно знать каждому историку – величина средней арифметической почти всегда виртуальна, ее значение может не встречается ни у одного объекта в изучаемой совокупности. К тому же,

это ей позволяет принимать любые дробные значения даже при дискретных признаках. Будь внимателен при интерпретации результатов. В истории все признаки весьма жестко подразделяются на непрерывные (способные принимать дробные значения) и дискретные, т.е. целые.

Средняя арифметическая бывает простая и взвешенная, которая вычисляется, если значения признака в изучаемой совокупности встречаются по несколько раз.

Проверь себя: Простую или взвешенную среднюю арифметическую величину можно определить по признакам Таблиц 1 и 24 ?

Таблица 24

Распределение крестьянских семей по количеству едоков и площади посева*.

Кол-во едоков	3.24	4.33	5.32	6.33	7.40	9.08
Площ. посева	2.34	3.25	4.36	5.14	6.10	7.18
Кол-во крест. семей	1248	1856	2312	3271	2803	1017

* Налоговые сводки НКФ СССР.- С.510.

А теперь поработай самостоятельно:

1. Определи величину средней арифметической по материалам таблиц 4, 15-18.

Подсказка : При интервальной разбивке признака в качестве его переменных значений берутся середины интервалов (допустим интервал «4 – 10», его середина определяется полусуммой крайних значений, т.е. $(4+10)/2=7$).

2. Схожесть величины средней арифметической в разных совокупностях еще не дает тебе право делать вывод о схожести самих совокупностей. У изучаемых объектов может быть разный размах вариаций признаков, разная амплитуда колебания. По данным таблиц 3, 14-17 посчитай величину R (размах вариации) и σ (среднего квадратического отклонения).

Историку всегда интереснее приблизиться к исторической действительности, чем ограничиваться общими рассуждениями о тенденциях развития, опираясь на виртуальные величины. В этом нам поможет Мода.

Подсказка: Мо – величина наиболее часто встречающегося значения признака в изучаемой совокупности или наиболее типичное среднее значение. На нее не влияют минимальное и максимальное значения признака, но ее точность повышается, если интервалы, по которым сгруппированы значения признака, примерно равны.

В дискретном ряду Мо определяется по наибольшей частоте и не требует вычислений. Если вдруг в твоей группировке равное максимальное значение частоты встречается более одного раза (би- или мульти-модальность), значит надо проверить правильность составления группировки, скорее всего она неоднородна.

В интервальном ряду величина Мо определяется по формуле.

А теперь поработай самостоятельно:

1. Определи величину Мо по материалам тех же таблиц, по которым ты вычислял среднюю арифметическую.

2. Сравни значения Мо и средней арифметической, полученные по одним и тем же данным. Подумай, почему они так отличаются? Какая из этих величин более достоверна, более близка к исторической реальности? Почему?

3. Вспомни, что приблизительное значение Мо можно определить графическим методом по гистограмме распределения. По имеющимся у тебя гистограммам определи Мо. Сравни ее графическое значение со значением, полученным по формуле – чем меньше разница, тем точнее ты построил график.

* * *

Следующая средняя величина – Медиана.

Подсказка: Ме – величина, определяющая значение признака, находящееся в середине изучаемой совокупности. Медиана делит всю совокупность пополам, так что число объектов, имеющих значения признака большее или меньшее, чем Ме, одинаково.

По своей математико-статистической природе Ме является самой представительной средней величиной. Она дает наиболее точные, досто-

верные результаты при больших колебаниях в значениях признака, в случае, если не определены крайние его значения.

И для дискретного, и для интервального ряда M_e придется вычислять. Начинать надо с построения ряда накопленных частот. Ты уже умеешь это делать, вспомни, как мы учились строить кумуляту...

Твой следующий шаг – определение суммы всех частот (это последняя величина в ряду накопленных частот – подсказываю, а то ты можешь начать все считать по новой). Вот это значение надо поделить на 2 и посмотреть – какая из накопленных частот впервые превысила полученную величину. Для дискретного ряда на этом завершаются все подсчеты, т.к. то значение признака, которое соответствует данной накопленной частоте и есть Медиана. Что касается интервального ряда, то далее в дело вступает формула и по ней продолжают вычисления.

Проверь себя. Давай вместе определим M_e по данным табл. 24 для признака «количество едоков».

Итак, ряд накопленных частот: 1248; 3104(1248+1856); 5416(1248+1856+2312); 8687 (1248+1856+2312+3271); 11490(1248+1856+2312+3271+2803); 12507 (1248+1856+2312+3271+2803+1017).

Теперь определим полусумму всех частот: $12507 : 2 = 6253,5$ и посмотрим, какая частота из ряда накопленных впервые превысила это значение. Это частота 8687, четвертая в построенном ряду, следовательно, четвертое значение признака в упорядоченной изучаемой совокупности соответствует Медиане. Это значение 6,33.

А теперь поработай самостоятельно:

1. По данным тех же таблиц, по которым ты вычислял среднюю арифметическую и M_o определи значение Медианы.

2. Сравни полученные значения Медианы со значениями M_o и средней арифметической, полученные по одним и тем же данным. Подумай, почему они так отличаются? Какая из этих величин более достоверна в данной группировке, более близка к исторической реальности? Почему?

3. Вспомни, что приблизительное значение Медианы можно определить графическим методом по кумуляте. По имеющимся у тебя кумулятам определи M_e . Сравни ее графическое значение со значением, полученным по формуле, и помни, что чем меньше разница, тем точнее ты построил график.

* * *

Для динамических рядов имеются свои средние показатели, также выступающие характеристиками тенденции развития.

Средний уровень динамического ряда определяется как простая средняя арифметическая уровней за равные промежутки времени. Для интервального – как средневзвешенная. Посмотри, по данным таблицы 10 можно будет подсчитать простую среднюю арифметическую, а по данным таблицы 11 – средневзвешенную.

Проверь себя. Попробуем определить среднегодовую общую площадь посева (средний уровень ряда) в Тираспольском уезде за период с 1914 по 1919 гг. по сведениям таблицы 10. Для этого используем формулу простой средней арифметической, поскольку временные промежутки в таблице равны.

$$\bar{Y} = (474,59 + 453,56 + 439,40 + 397,11 + 422,0 + 331,50) / 6 = 419,69$$

А теперь определим прогнозировавшуюся среднюю продолжительность жизни мужчин в СССР в 1959-1979 гг. Используем для этого данные таблицы 11 и формулу средневзвешенной, т.к. мы имеем дело в исходных показателях с интервальным рядом и временные промежутки здесь не равные.

$$\bar{Y} = (64,4 \times 2 + 65,7 \times 9 + 64,5 \times 6 + 62,9 \times 3 + 62,5 \times 1) / 21 = 64,68$$

Обрати внимание, что хронологические промежутки считаются буквально по пальцам, вычитание из большей величины меньшего значения здесь не подойдет. С 1961 года к 1969 году прошло не $1969 - 1961 = 8$, а 9 лет, потому что крайние даты включаются в изучаемый промежуток времени.

А теперь поработай самостоятельно:

1. Рассчитай средние показатели для площади посева зерновых (табл. 10) и для ожидаемой продолжительности жизни женщин в СССР (табл. 11).
2. Для каких динамических рядов можно вычислить среднюю хронологическую? Найди в «Задачнике» подходящую таблицу и определи среднюю хронологическую величину.

3. По данным таблицы 22 подсчитайте средний абсолютный прирост показателей.

4. По данным таблиц 10, 22 и 23 определите средний темп динамики и средний темп прироста. Подсказка: При этом будь особенно внимательным к тенденции изменения показателей. Дело в том, что уровни ряда за изучаемый промежуток времени могут возрастать, сокращаться, вновь возрастать и игнорировать эти колебания нельзя, т.к. тем самым ты «затушевываешь» исторический процесс. Вот посмотри, для примера, на показатели смертности, приведенные в таблице 12. С ноября 1921г. по март 1922 г. уровень смертности неуклонно возрастает, затем начинает снижаться. Средняя скорость этих процессов разная, процессы противоположной динамики (рост-сокращение). Однако если сравнить только крайние показатели, то можно констатировать только рост уровня смертности. Следовательно, в данном случае надо вычислять два средних темпа динамики и два средних темпа прироста. При определении того, сколько темпов придется вычислять для того или иного показателя, полезно обратиться к графическому представлению данных.

Напоминаю, хотя ты уже сам в этом убедился (конечно, если выполнил все задания), что средние показатели динамических рядов особенно полезно вычислять при сравнительном анализе двух и более признаков.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Очень заманчиво в историческом исследовании получить неопровержимые данные, выводы и оценки. Определенную помощь в этом может оказать метод корреляционного анализа, с помощью которого исследователь вносит количественную определенность, выясняя степень влияния одного признака на другой (парная корреляция) или на несколько признаков (множественная корреляция).

Подсказка: 1. Выбор корреляционного коэффициента зависит от характера исходных данных, а также от конечной цели применения корреляционного анализа.

2. Не зависимо от выбранного вида коэффициента корреляции, его значение определяется и интерпретируется в одних и тех же пределах – от -1 до $+1$.

3. Отрицательное значение коэффициента говорит об обратной связи между рассматриваемыми признаками, положительное – о прямой. Не забудь, что при изменении в значении одного признака меняется не второй признак, а его среднее значение.

4. У каждого коэффициента корреляции есть свои доверительные интервалы, т.е. величины, помогающие качественно оценить количественную величину, числовое значение коэффициента.

5. Всегда помни, что не всякие признаки, подпадающие под корреляционный анализ, обладают историческим смыслом. Прежде чем начнешь считать – подумай, надо ли это делать.

Проверь себя. 1. Вспомни названия коэффициентов корреляции, обозначаемых следующими символами: r , Q , K_k , p , k_s .

2. По какому коэффициенту можно определить, на сколько процентов результативный признак меняется под воздействием факторного? Как подсчитывается коэффициент детерминации?

3. Что можно определить по автокорреляции? Для каких группировок применяется этот показатель?

А теперь поработай самостоятельно:

Подсчитай возможные коэффициенты корреляции по данным приведенных таблиц:

Таблица 25

Распределение доходов и стоимости рабочего скота
в крестьянском хозяйстве (в рублях).*

До- ход	648	803	513	706	541	230	724	1461	828	1437	973
Ст. ско та	353	51	6	273	100	22	287	158	189	570	163

* Таблица учебная. Данные условные.

Таблица 26

Успеваемость студентов-геологов МГУ в 1966/67 уч. году*

Успеваемость	Кол-во студ-в	Мужчин	Женщин
Отлично	12	5	7
Хорошо	28	18	10
Удовлетворительно	25	14	11
Неуспевающих	9	4	5

* Таблица учебная. Данные условные.

Таблица 27

Распределение сельского населения Среднего Поволжья по полу и грамотности.*

Население	Грамотное	Неграмотное
Мужчины	2004694	2254465
Женщины	988684	3849615

*ЦСУ СССР. Всесоюзная перепись населения 1926 г. – Т. 3. – С. 206–207.

Таблица 28

Распределение едоков и посевов в крестьянском хозяйстве Волжско-Камского края по обеспеченностью крупным скотом (в %).*

Крупн. скот	Без скота	1	2	3	4	5 и более
Кол-во едоков	3.24	4.33	5.32	6.33	7.40	9.08
Площ. посева	2.34	3.25	4.36	5.14	6.10	7.18

* Налоговые сводки НКФ СССР.- С.510.

Таблица 29

Размер дореформенного надела крестьян Симбирской губернии.*

Надел в дес.	0-2	2-3	3-5	5-7	7 и более
Барщин. в % %	1.8	18.4	63.5	15.2	1.1
Оброчн. в %%	12.4	17.5	48.2	13.3	8.6

* Канатов В.И. Своеобразие проведения реформы 1861 г. в Среднем Поволжье // Вестник МГУ, серия 9, 1964, №8.

Таблица 30

Распределение лишенцев по времени поражения в правах.*

Год	1924	1925	1927	1928	1929	1930	1931	1932	1933	1934
Число лишений	2	10	3	3	2	12	5	5	5	36

* Социальный портрет лишенца (на материалах Урала). – Екатеринбург, 1996. – С.206.

Таблица 31

Распределение лишенцев по времени и мотиву поражения в правах.*

Мотив	1929	1930	1931	1932	1933	1934	1935
Кулак	57	132	66	9	39	28	0
Эксплуататор	62	31	47	10	0	17	19
Торговец	16	18	21	24	13	18	8
Священник	5	4	7	11	6	9	7
Родство с лишенцем	27	29	31	29	25	33	17
Служба в белой армии	12	7	16	8	12	11	9
Другие	11	1	14	6	10	7	13

* Таблица учебная. Данные условные.

Таблица 32

Распределение служащих г.Казани по возрасту и семейному состоянию.*

Возраст	Холост.	Семейн.	Вдовы	Разведен.	Не указ.
15-19	379	13	0	2	0
20-24	1105	696	2	11	0
25-29	672	2212	12	28	10
20-34	240	2303	19	23	6
35-39	116	1932	14	23	2
40-44	81	1511	24	17	3
45-49	49	1073	33	14	-
50-54	22	757	51	8	2
55-59	28	523	39	1	1
60-64	11	232	28	1	-
65-69	6	119	23	1	-
70 и старше	1	22	11	-	-
Не указ.	1	-	-	-	-
Итого	2711	11393	256	129	24

* ЦСУ СССР. Всесоюзная перепись населения 1926 г. – Т.20. – С. 133.

КОНТЕНТ-АНАЛИЗ

(Из истории метода)

Контент-анализ (contents – англ.: содержание, содержимое) – научный метод, предполагающий формализованное исследование содержания текстовых массивов в целях выявления и измерения, представленных в них социальных, культурных, ментальных особенностей. Контент-анализ применяется при изучении источников, инвариантных по структуре или существу содержания, но внешне бытующих, как несистематизированный, беспорядочно организованный текстовой массив. Суть контент-анализа, как

исследовательского метода, состоит в восхождении от многообразия текстового материала к абстрактной модели содержания текста. В указанном смысле, контент-анализ является одной из номотетических процедур, используемых в сфере применения идеографических методов.

Термин *content-analysis* впервые начал применяться в конце XIX – нач. XX вв. в американской журналистике (см. работы Б. Мэттью, А. Тенни, Д. Спид, Д. Уипкинс). У истоков методологии контент-анализа находились американский социолог Г. Лассуэл и французский журналист Ж. Кайзер.

Так, в начале 60-х гг. Г. Лассуэл осуществил попытку политологического анализа СМИ, исходя из учета формальных критериев. Он ввел в научный оборот некую абстрактную единицу: «слово». Целью работы Лассуэла было получение собственно социологического результата на нетипичном для социологии материале: текстах печатных изданий. Исследователь проделал огромную работу, но, поскольку в методике Лассуэла качественные оценки не были адекватно соотнесены с количественными методами, результаты его трудов с трудом поддавались верификации.

В начале 60-х гг. Ж. Кайзер разработал оригинальную методику статистического анализа периодики. В ее основе лежал подход к тестовому массиву, как информационной системе. Тем самым Кайзер сформулировал теоретическую базу последующего распространения социологических методов в сферы изучения всех нарративных источников, включая эпиграфический и эпистолярный материал.

В работе Ж. Кайзера акцентировалось внимание на внешней форме организации материала: его расположении, оглавлении, оформлении и т.д. Кайзер разработал целый комплекс исследовательских процедур, обеспечивающих полную формализацию, как единичного газетного номера, так и совокупности однотипных периодических изданий. Тем самым Ж. Кайзер сформулировал систему, позволяющую фиксировать развитие тенденций в публикациях СМИ.

Свое дальнейшее развитие «кайзеровское направление» методологии контент-анализа получило в работах Э. Морэн. Э. Морен ввела в научный оборот термин «единица информации» – семантический блок, содержание которого отвечает на вопрос: «О чем говорить?» Последнее обстоятельство сделало возможным изучение любых форм организации текстового материала, причем, как на терминологическом уровне, так и на уровне фразы, абзаца, статьи и даже целых книг. Тем самым, Э. Морэн разрушила

критерий однородности, применявшийся ранее при статистической обработке нарративов. Взамен, она предложила идеологию «семантических групп», которые, по ее мнению, должны учитываться по тематическому признаку. Кроме того, Э.Морэн разработала концепцию «тоню» материала, который определялся социометрически: «положительная информация», «отрицательная», «нейтральная».

В отечественной историографии методика контент-анализа описывается, как совокупность (комплекс) исследовательских процедур, направленных от формирования абстрактного объекта исследования и формулирования его конкретных задач, до интерпретации результатов и их статистической обработки.

В источниковедческих исследованиях контент-анализ используется не менее трех десятков лет. Суть применения метода состоит в формализации источникового материала.

Адекватность любого источника есть, по существу, его потенция к субъективации. Абстрагирование в процессе формализации, опосредовано содержанием сознания исследователя и потому, с неизбежностью, также оказывается субъективацией. Отсюда, формализация источника, *ipso facto*, оказывается созданием нового источника, т.е. творческим актом.

В социологии контент-анализ позволяет отслеживать тенденции развития социальных процессов, с целью прогноза функционирования социальных систем или их компонентов. В источниковедении этот метод выступает средством фундирования суждений о содержании источников.

В отечественной историографии контент-анализ, как правило, применялся для выявления тенденций трансформации взглядов конкретных авторов во временном аспекте, либо при сравнении позиций различных авторов по одному и тому же вопросу.

В силу известных экономических, социальных и политических причин, с конца 90-х гг., методический арсенал исторической науки практически не пополнялся. Вот почему вопросы методологии контент-анализа как метода исторических исследований, и теперь сохраняют свою актуальность.

Описывая процедуру контент-анализа, можно выделить несколько этапов, а именно:

Начальный этап исследования: На этом этапе, как правило, формулируется т.н. «эмпирическая теория» исследования. То есть, в ходе подготовки к проведению контент-анализа, ученый, как правило, систематизирует гипотезы, существующие в контексте данной проблематики и «отсеивает» те из них, которые не поддаются верификации на данных обусловленного информационного массива.

Кроме того, на первом этапе такого исследования формируется идеология исследовательской программы. Иначе говоря, в этот момент формулируются философские (теоретические), предпосылки, определяются основные задачи исследования и, вытекающие из них, базовые правила процедуры, а также выявляется логическая последовательность операций по классификации всего корпуса существующих гипотез.

С методической точки зрения, именно на этом этапе происходит актуализация системного соотношения цели, объекта и средств исследования и определяется их оптимальная конфигурация в процессе предстоящей работы.

Операциональный этап исследования: На втором этапе исследования создается основа для верификации используемых дефиниций и терминов. Как правило, в этот момент в сплошном текстовом массиве выделяются относительно автономные части – «единицы контекста», релевантные, как содержанию всего текстового массива, так и общей совокупности сформулированных дефиниций и терминов.

При этом при выявлении латентной информации, представляется целесообразным использовать двухэтапную методику Р.Л. Лебедева-Р.В.Рывкина (см.: Лебедев Р.Л., Рывкин Р.В. О двух функциях категорий в контент-анализе // Методологические и методические проблемы контент-анализа. Тезисы рабочего совещания социологов. М.-Л., 1973, Вып.1., с 71), которая предполагает существование в тексте двух категорий анализа: категорий структурирования как средства извлечения латентной информации и категорий интерпретации, зависящих от первых.

Этап счета: На третьем этапе исследования проводится собственно *контент-анализ*. То есть, осуществляется сбор информации, учет индикаторов, первоначальная обработка результатов, включающая установление видов и содержания отношений связей между элементами счета (индикаторами контент-анализа). Причем, последнее – как внутри совокупности

данных, так и по отношению к самой совокупности как целостной информационной системе.

На этом этапе, принципиальную важность обретает то обстоятельство, что полученные данные должны гарантировать временную, интересубъективную и инструментальную стабильность исследования, а также в полной мере соответствовать (находится в контексте) его теоретическим посылам.

В случаях, если полученные данные расходятся с теоретическими посылами, применение контент-аналитических процедур завершается, и исследование начинается с начала.

Остаточный анализ данных. На четвертом этапе осуществляется агрегирование (интерпретация числового материала в целях его обобщения), структуризация и группировка полученных данных для формирования адекватных моделей в контексте задач, поставленных на первом этапе исследования.

В целом, нужно сказать, что анализ полученного материала напрямую зависит от трех режимов восприятия информации: режима «настройки», режима восприятия основной и режима восприятия возвратной информации. Причем, если первые два режима важны на первом и втором этапах контент-анализа, то третий режим приобретает особое значение на четвертом этапе и оказывается полностью обусловленным качеством работ на этапах 1-3.

Отдельно следует остановиться на проблеме отбора источникового материала.

Для того, чтобы применение контент-анализа было успешным, источник должен отвечать определенным требованиям. То есть, прежде всего, при выборе источника нужно определить, в какой мере его содержание соответствует поставленной задаче. Необходимо также изучить все существующие источники по данной проблеме и, если понадобится, выявить оптимальный размер репрезентативной случайной выборки.

Обычно выделяют три уровня стратегии исследовательского поиска при обработке текстовых массивов: формулятивный, описательный и экспериментальный планы. В исторических исследованиях чаще всего используют описательный план, целью которого является системное качественно – количественное описание источника.

Описательный план всегда содержит элементы экспериментального. Сбор информации в контексте данного плана осуществляется на основе тотального или выборочного обследования (с обязательными расчетами размеров допустимой выборки). Описательный план предполагает особую строгость интерпретации понятий и точность в регистрации данных. Исследование по описательному плану обычно завершается классификацией данных, характеризующих структуру объекта исследования. К сказанному следует добавить, что изначально, описательный план, как правило, готовится в виде наброска, который уточняется и дополняется по мере «вживания» историка в источниковый материал и при реализации последовательности процедурных шагов. Принципиальное значение для реализации описательного плана контент-анализа имеет технический инструмент историка.

В заключение необходимо сказать, что контент-анализ является весьма продуктивным методом исторических исследований, позволяющим порой получать весьма неожиданные результаты при изучении, как отдельных нарративных источников, так и больших несистематизированных информационных массивов.

Пример использования КОНТЕНТ-АНАЛИЗА в исторической науке.

Архангельская И.Д. К вопросу изучения периодической печати методами контент-анализа / отв. Ред. Косс Б.М. // Методы количественного анализа текстов нарративных источников. – М., 1983. – С. 110-125.

И.Д. Архангельская в своей статье на материалах «Торгово-Промышленной газеты» предприняла опыт разработки методики источниковедческого изучения периодики. Хронологические рамки исследования – 1910–1914 гг.

Исследование носило характер эксперимента. Свою задачу автор видел в разработке общих принципов и технических приемов выявления информативных возможностей газеты, отборе и систематизации как определяющих, так и второстепенных проблем, классификации информации по типам, ее концентрации, т.е. формализации информации, подготовке ее к оцениванию. В итоге полученные результаты могут классифицироваться как предварительные. Предложенная автором методика является перспек-

тивной в плане предварительного анализа ежедневной газеты. Подобная обработка газетной информации рассматривается как необходимый подготовительный этап для использования контент-анализа. Метод ориентирует на теоретическое осмысление изучаемого объекта.

При разработке методики извлечения и обработки информации И.Д.Архангельская ограничивалась двумя основными структурными подразделениями – отделами: Передовые и Хроника. Передовые статьи являются лицом издания. Им присущ аналитический характер. Данные статьи являются концентрированным выражением информационно-справочного характера «Т. – П.Г.». Они, с одной стороны, питают отдел «Передовые» материалом (однако, чаще косвенно, чем прямо), а с другой – являются резервуаром различных мнений в ходе обсуждения какой-либо проблемы, отражением событий общественной жизни. В этом автор увидел их основную функцию. Особенности этих статей наложили отпечаток и на характер пометок «Ред.»

Автор разделил информацию на 3 группы: А – «сквозная», В – «механическая», С – «сплошная». Группа А охватывает период 1910-1911 гг. В рамках каждого года рассматривалась информация первой недели декабря. Группа В охватывает 1912 г. и 1914 г. Рассмотрению подвергался каждый 10-й номер газеты. Группа С на ограниченном временном отрезке – октябрь-декабрь 1913 г. Применяя выборочный метод, автор исходил из гипотезы, что, поскольку издание является ежедневным, массовым по своему типу и достаточно четко организованным, внутри него существуют основные акценты, которые могут быть выявлены посредством выборки, примененной на большом временном отрезке (1910-1914 гг.). Автор предполагал, что информация конца года носит преимущественно оценочный характер, аккумулируя нерешенные проблемы и выдвигая новые. Элементы формализации, заложенные уже в самом отборе газетного материала, получили свое дальнейшее развитие в его проблемной систематизации, классификации информации по типам, выявлении ее внутренних взаимосвязей. Работа велась в 4 этапа:

I этап

1. Составление предварительного перечня проблем
2. Накопление материала
3. Изучение внутренней организации источника, общая характеристика его структурных подразделений, выявление структурных взаимосвязей и распределение информационной нагрузки, введение первичной кар-

точки учета информации и авторской карточки, формирование проблемной и авторской картотеки.

II этап

1. Систематизация и концентрация информации:

- Корректировка перечня проблем на основе содержательного анализа
- Обработка первичной карточки информации, уточнение проблематики, выявление внутренних и хронологических взаимосвязей, проблемно-сопоставительный анализ

2. Систематизация информации по типу и содержанию:

- Выявление ключевых элементов зафиксированной информации
- Определение типа информации
- Общая оценка содержания информации
- Проверка карточки на соответствие номерным индексом

3. Концентрация информации:

- Групповой уровень
- Межгрупповой уровень

III этап

1. Анализ информации во внутренних взаимосвязях

- Выявление главных проблемно-координационных центров и линий взаимосвязи между ними и внутри них
- Проблемно-комплексная характеристика источника

IV этап

1. Оценка информации

- Выявление характерных черт издания с точки зрения его классовой направленности
- Проблемно-комплексный сравнительный анализ информации (с привлечением других изданий)

В качестве «организатора памяти» на первом этапе формализации автором предлагается первичная карточка учета информации, в которой зафиксированы следующие категории анализа:

№№ п/п		+
Число, месяц, год		
Страница		
Отдел, рубрика		
Название информации		
Автор		X(XX)

где + – буквенный индекс группы, в рамках которой ведется исследование

X – номерной индекс проблемы, к которой относится карточка

XX – номерной индекс проблемы, которая тесно связана с первой.

Карточками такого вида укомплектовывается проблемная картотека, указателем к которой является перечень проблем. Составляя перечень, автор брал сведения из рубрик газет, а также изучения литературы по вопросам истории экономической политики и экономического развития России

Особые требования в процессе формализовано-системного анализа необходимо предъявить к такой категории как «краткое содержание статьи». Именно эта категория оказывает определяющее влияние на процесс систематизации информации по типу и содержанию в рамках второго этапа формализации информации. Во-первых, необходимо строго сохранять структуру статьи как костяк мысли ее автора, а, во-вторых, точно выделять главные ее элементы. Картотека дала возможность частично снять «дублирующие» проблемы. Вместе с тем работа с ней показала, что интуитивный подход к комплектованию информации по проблемам нуждается в подкреплении более устойчивым и менее субъективными принципами. Здесь на помощь приходят методы контент-анализа.

Одновременно с проблемной происходило комплектование авторской картотеки, основой которой была карточка следующего вида:

Подпись	Частота публикаций	Тематика публикаций	Сведения об авторе
	X – (XX)		

где X – год публикации

XX – число опубликованных в этом году статей и заметок.

Вопрос об авторстве представляет наибольшие трудности при изучении газеты. Одним из путей его решения на взгляд автора систематизация данных, содержащихся в самом источнике.

Отправным моментом второго этапа формализации информации является обработка первичной карточки учета информации. Приведем пример обработки первичной карточки:

№№ п/п	№70	Заполняется краткое содержание информации 5 (9)
Число, месяц, год	23.3.1912 г.	
Стр.	2	
Отдел, рубрика	Хроника, пути сообщения	
Название информации	Экстренное собрание пароходовладельцев	
Автор	-	

Ключевые элементы зафиксированной информации:

– о невозможности угольных фирм Петербурга выполнить заказы в связи с английской забастовкой и об использовании запасов топлива владельцев пароходных судов; предложения: о закупке более дорогого германского угля, об использовании донецкого угля (не приняты);

– решение: ограничиться пока наличными запасами. Тип информации – экстренная. Оценка: 1. причина состоит в сильной зависимости России от поставки английского угля; 2. несмотря на экстренный характер собрания, решение не найдено. Вынесенное решение: «Крупным судовладельцам поддержать более мелких ввиду общности интересов навигации». В итоге:

а) Некоторые крупные судовладельцы отнюдь не бедствуют, несмотря на английскую забастовку;

б) Сама апелляция к «общности интересов навигации» звучит в подобных условиях капиталистической действительности как насмешка.

Номерной индекс 5 соответствует проблеме «Ввоз и вывоз сырья», а также готовой продукции». Номерной индекс (9) – сопряженной с ней проблеме, а именно «Кризис топлива (топливный «голод»))».

В результате проведенного эксперимента удалось, прежде всего, определить степень эффективности применения каждого из трех принципов выборки, выявить предварительные проблемно-информационные возможности источника и наметить внутренние линии взаимосвязи информации. Опыт показал, что применение к источнику «сквозного» принципа выборки обеспечило выход информации в 6 проблем, «механического» – в 10, «сплошного» – в 23. Таким образом, налицо оказалось преимущество сплошного принципа изучения материала даже на ограниченном временном отрезке.

В результате проведенного сопоставительного анализа на межгрупповом уровне из всего комплекса проблем были выделены две, характеризующие в целом период 1910-1914 гг.:

1. Статьи по крупным проблемам экономического развития ведущих капиталистических государств: в т.ч. и России
2. Состояние финансов России, финансовые проблемы

Выделение двух проблемно-координационных центров, является объективным отражением наличия взаимосвязей аналитического информационного газетного материала и позволяет характеризовать «Т.-П.Г.» как комплексный источник. Однако выводы, полученные экспериментальным путем, требуют подтверждения более точными количественными методами.

Предложенная методика источниковедческого изучения периодики на примере «Торгово-Промышленной Газеты» призвана, во-первых, помочь историку получить самое общее представление о данном печатном органе. Во-вторых, переориентировать историка с иллюстративного на комплексный подход к решению конкретной проблемы по материалам периодической печати.

* * *

На этом мы завершаем наш «Задачник». Надеюсь, он поможет любому интересующемуся разобраться в учебном курсе «Применение математических методов в историческом исследовании». Желаю успеха на практических (лабораторных) занятиях и в дальнейшей работе.

СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ:

1. Учебно-методическая литература

- 1.1 Информационные технологии для гуманитариев. – Москва-Саранск, 1998. – 215 с.
- 1.2 Количественные методы в исторических исследованиях. – М.: Высшая школа, 1984 – 384 с.
- 1.3 Компьютеризированный статистический анализ для историков. – М.: изд-во МГУ, 1999. – 186 с.
- 1.4 Федорова Н.А. Математические методы в историческом исследовании. – Казань, изд-во Форт-Диалог, 1996. – 110 с.

2. Специальная литература

- 2.1 Измайлова М.О., Рахманкулов И.Ш. Категория «средняя величина» и ее методологическое значение в научном исследовании. – Казань, изд-во КГУ, 1982. – 145 с.
- 2.2 Математические методы в социально-экономических и археологических исследованиях. – М.: Наука, 1981. – 415 с.
- 2.3 Миронов Б.Н. История в цифрах. – Л. : Наука, 1991.- 178 с.
- 2.4 Миронов Б.Н., Степанов З.В. Историк и математика. – М.: Наука – 165 с.
- 2.5 Миронов Б.Н. Историк и социология. – Л.: Наука. – 1984. – 175 с.
- 2.6 Славко Т.И. Математико-статистические методы в исторических исследованиях. – М.: Наука, 1981. – 158 с.
- 2.7 ЭВМ и математические методы в исторических исследованиях. М., 1993. – 204 с.
- 2.8 ЭВМ и математические методы в исторических исследованиях. – М., 1994. – 223 с.

3 Периодические издания

- 3.1 Информационный бюллетень ассоциации «История и компьютер»/
Издается с 1994 года.
- 3.2 Круг идей. Сборник статей участников периодических конферен-
ций ассоциации «История и компьютер»/ Издаётся с 1994 г.